Potentials of Linking Administrative Data and Survey Data for Inequality Research

Rudolf Farys, Oliver Hümbelin, and Ben Jann

University of Bern and Bern University of Applied Sciences

BigSurv18
Barcelona, October 25-27, 2018

Contents

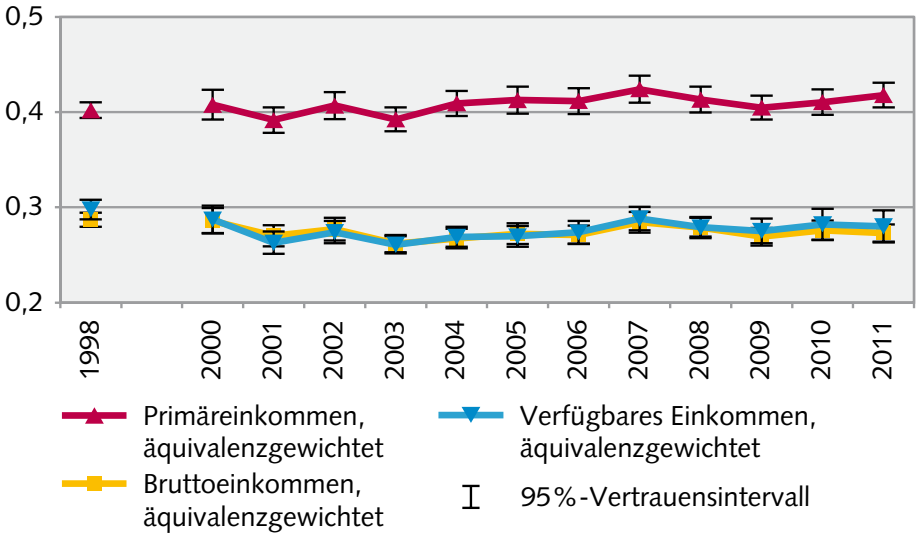
1. Introduction
2. Data linkage
3. Some preliminary results
4. Reweighting
5. Conclusions

Inequality in Switzerland

- Income inequality increased in many OECD countries (OECD 2008, 2011, 2015; Salverda et al. 2014) .
- Switzerland: Results are ambiguous, depending on data source
  - Survey-based estimates indicate stable, or even decreasing inequality (e.g. Household Budget Survey).
  - Tax-data-based estimates indicate increasing inequality, in particular at the upper end of the distribution.

Inequality in Switzerland

Entwicklung der Gini-Koeffizienten 1998 bis 2011, Gesamtbevölkerung[1] **G 6**

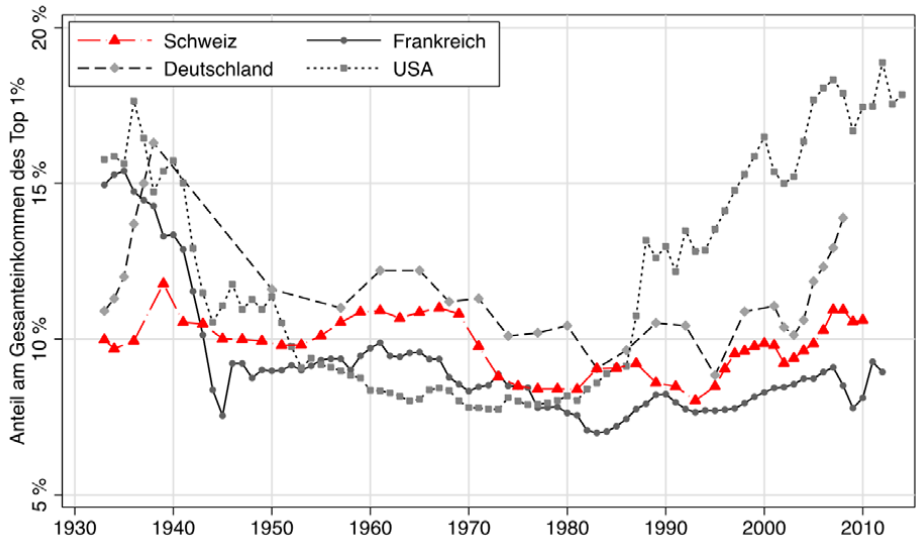


[1] Berechnungen einschliesslich der negativen Einkommen

Quelle: Haushaltsbudgeterhebung © *BFS*

(Bundesamt für Statistik 2013)

Grafik 1. Einkommensanteil des Top 1% im internationalen Vergleich



(Martinez 2017)

Survey data vs. tax data

- Survey data:
  - Pros:
    - ★ measurement of income based on theory-guided income definitions
  - Cons:
    - ★ strong middle-class bias, underrepresentation of the top and the bottom
    - ★ small sample sizes
- Tax data:
  - Pros:
    - ★ full census
    - ★ income and assets in great detail
  - Cons:
    - ★ measurement of income for administrative purposes; some components lacking (e.g. social assistance)
    - ★ tax subjects, not households
    - ★ little additional information
    - ★ difficult to obtain from all cantons of Switzerland

Goal our project

- Obtain better data for inequality analysis in Switzerland by linking
  - various administrative data sources at the cantonal or federal level containing information on income and wealth (e.g. tax data, social security data),
  - official population registry to define the population and identify households,
  - large-scale survey data (e.g. Swiss structural surveys).
- Important restriction
  - Tax data are owned by the cantons; not all cantons agree to (or capable of) providing data.
  - We therefore also evaluate whether data based on selected cantons can be used to estimate the level of inequality in whole Switzerland.

1 Introduction

2 Data linkage

3 Some preliminary results

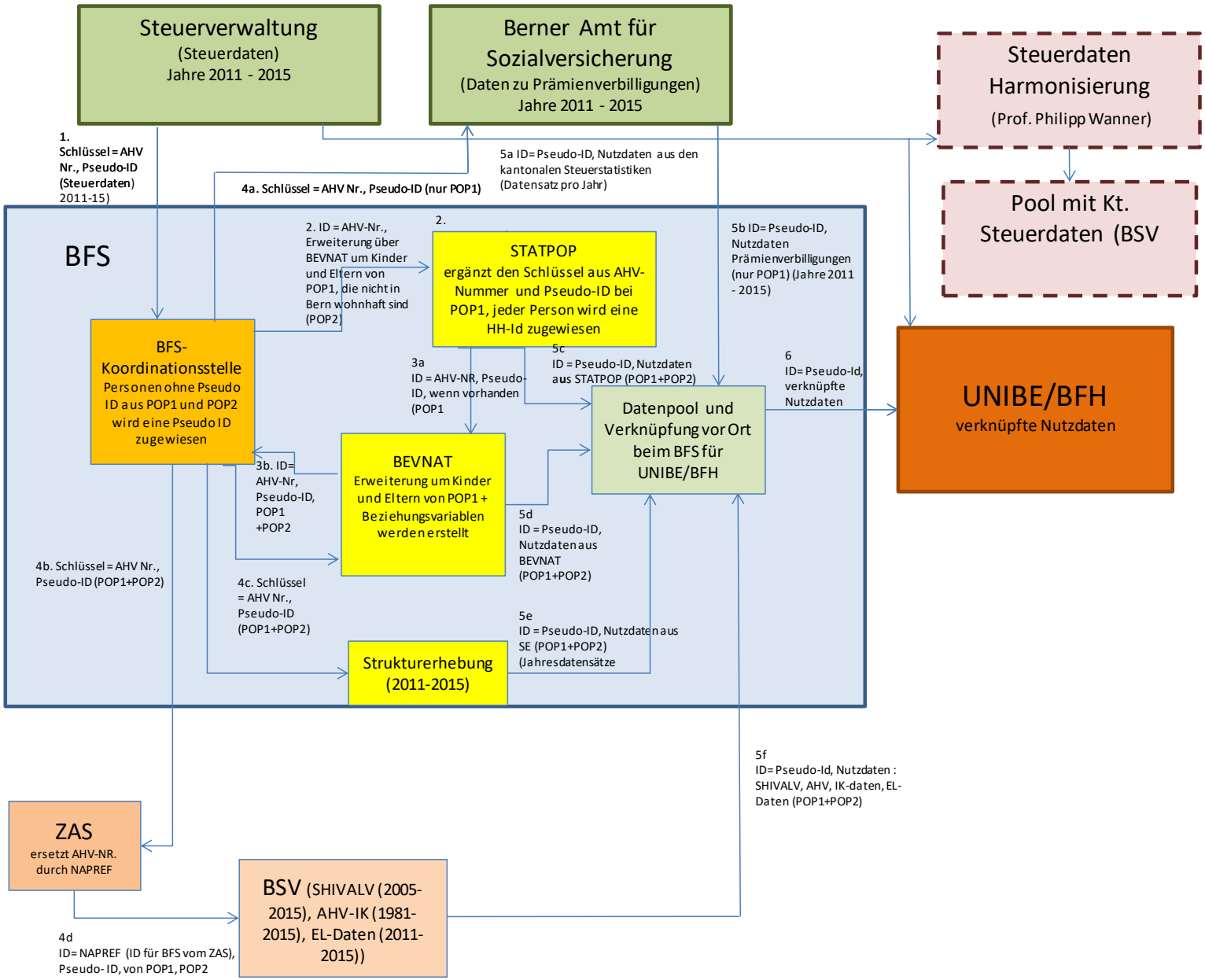
4 Reweighting

5 Conclusions

Data linkage

- Data from various sources are linked based on social security numbers, which are available in many sources since around 2010.
- Key goals:
  - Describe financial situation of households as detailed and as complete as possible.
  - Obtain data for all members of a clearly defined population.
  - Be able to identify households.
  - Enrich the data with background information on households and the respondents based on available survey data.

Data linkage



Data linkage

- Financial data
  - Detailed individual-level tax data from cantonal tax authorities (starting in 2011).
  - Social assistance and other non-taxed transfers from the Federal Social Insurance Office.
  - Additional information on earnings and transfers from other sources such as cantonal social insurance offices.
- Population and households
  - STATPOP (population register) and BEVNAT (births, marriages etc.) maintained by the Swiss Federal Statistical Office (SFSO)
- Survey data
  - currently: Structural surveys by the SFSO (> 200'000 people/year)
  - many other possibilities ...
- Challenges
  - Many data owners involved, data protection issues and complicated contracts, lots of signatures, everything takes a long time ...
  - Data cleaning, consistency, etc.

1 Introduction

2 Data linkage

3 Some preliminary results

4 Reweighting

5 Conclusions

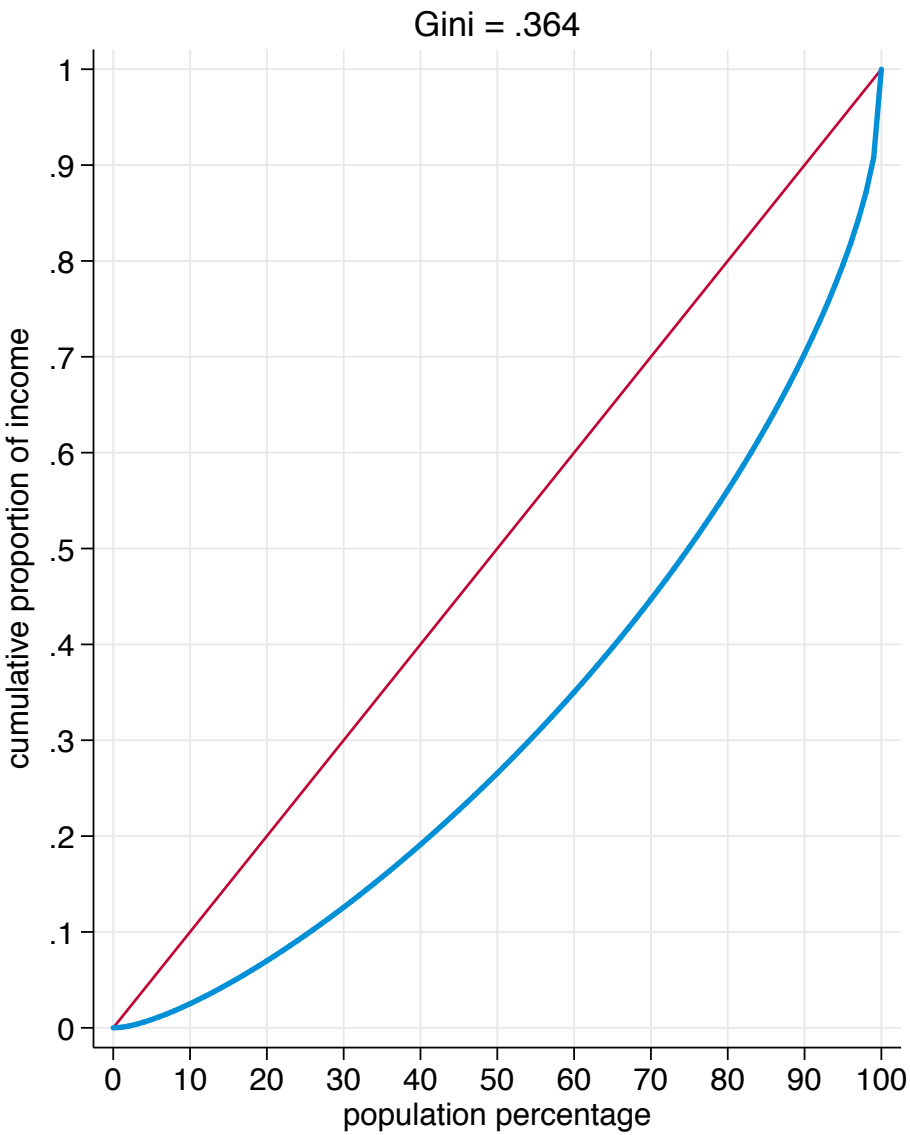
Data on canton of Bern

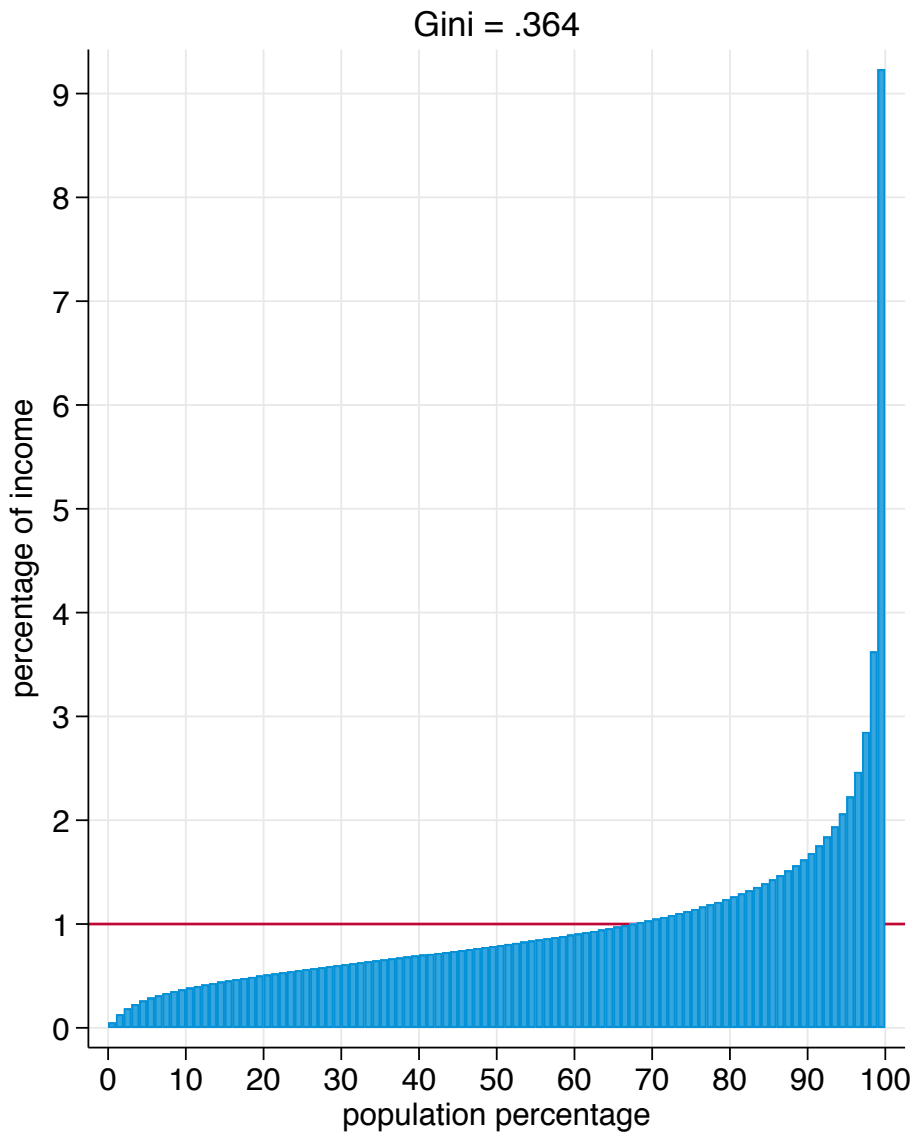
- We just received the first chunk of linked data, which covers the canton of Bern (about 1 million inhabitants; about 12% of the Swiss population).
- The following results are *very* preliminary. There was not much time yet to do careful data cleaning and sort out the details.
- We will show some descriptive results on the income distribution, compare there results to survey estimates, and then try to make some inference on Switzerland as a whole.

Outcome

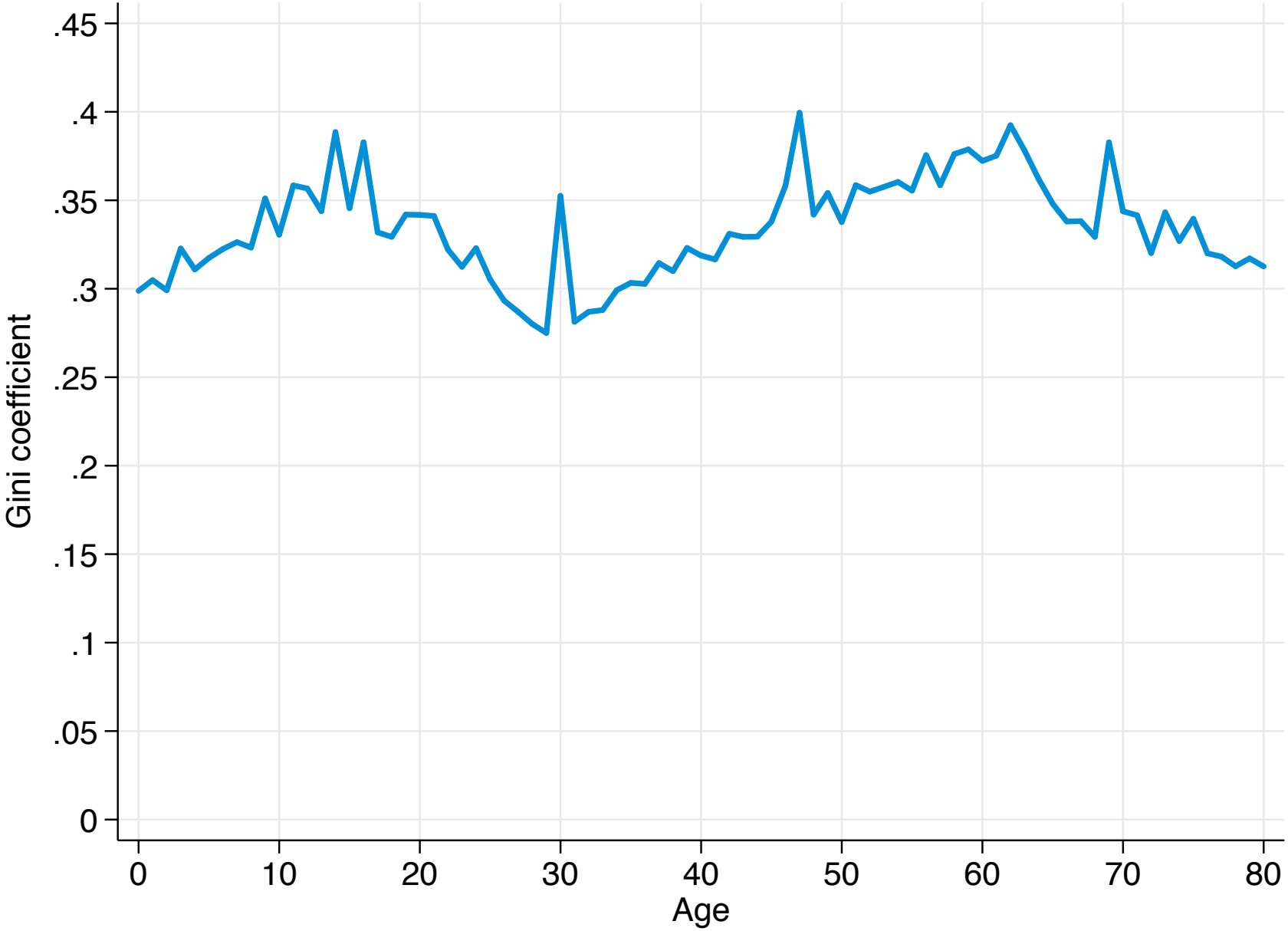
- Equivalized disposable income
  - earnings, capital income
  - pensions, social assistance and other transfers
  - minus social security contributions, health insurance, and taxes
  - equivalized using revised OECD scale
  - full population aged 0–109
  - canton of Bern
  - households containing adults without tax record excluded (mostly people taxed at the source)
  - in addition: households with negative or 0 income excluded (conservative strategy)

Overall Lorenz curve/percentile shares



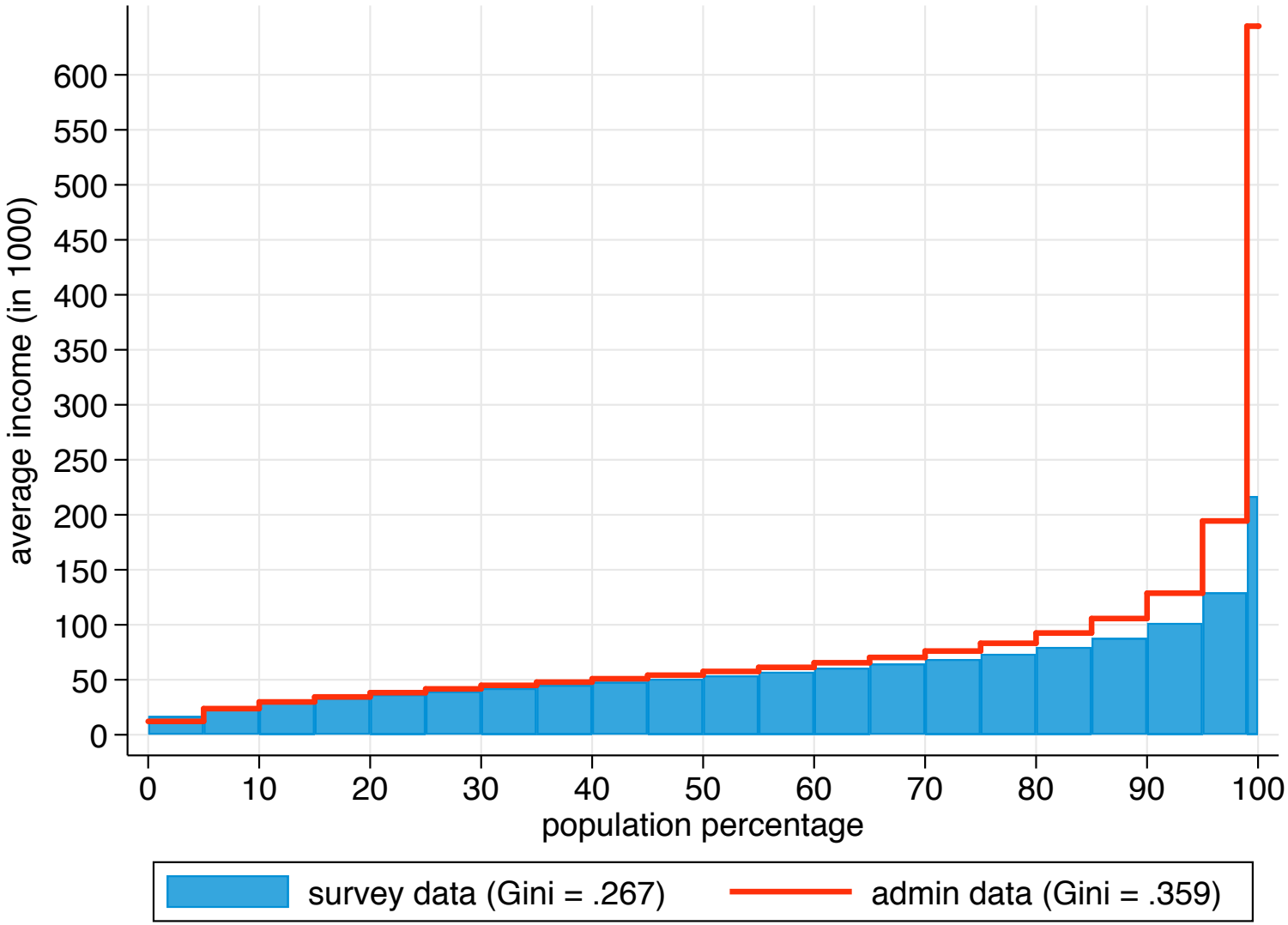


Gini by age


.45
.4
.35
.3
.25
.2
.15
.1
.05
0
Gini coefficient
0
10
20
30
40
50
60
70
80
Age

Comparison to survey data (adults only)

- The survey data is from the household budget survey by the SFSO that was used to compute the results in the graph on slide 4.



1 Introduction

2 Data linkage

3 Some preliminary results

4 Reweighting

5 Conclusions

Reweighting

- One limitation of our project is that we will not receive tax data from all cantons.
- We are thus looking into whether it is possible to reweight the cantonal data in a way such that it is representative for whole Switzerland.
- Available at the Swiss level are public tax statistics from the Federal Tax Administration (FTA).
  - Quite detailed, including mean, median, Gini coefficients of taxable income at the municipality level.
  - However several problems
    - inappropriate income measurement (taxable income)
    - statistics are for "tax subjects", not households or individuals
    - aggregate data

Idea

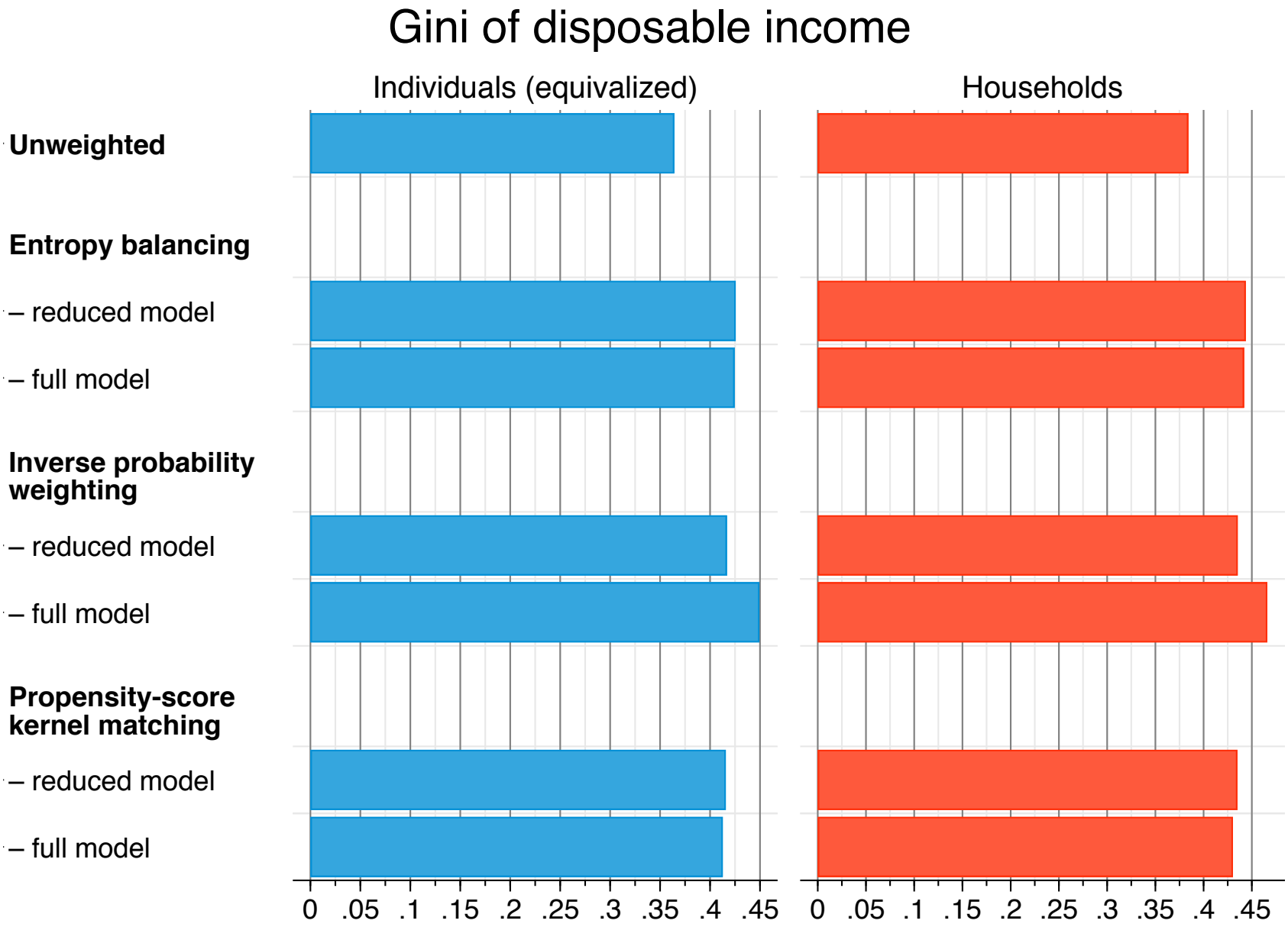
- Use the FTA indicators (as well as other indicators from the SFSO) to derive weights that can be applied to cantonal tax data.
- The weights are constructed at the municipality level. The goal is to reweight the municipalities of a canton such that they look like Switzerland.
- These weights can then be used when analyzing the individual-level tax data from the canton.
- The procedure should work if there is enough heterogeneity among municipalities and if strong predictors for inequality at the municipality level are available.

Method

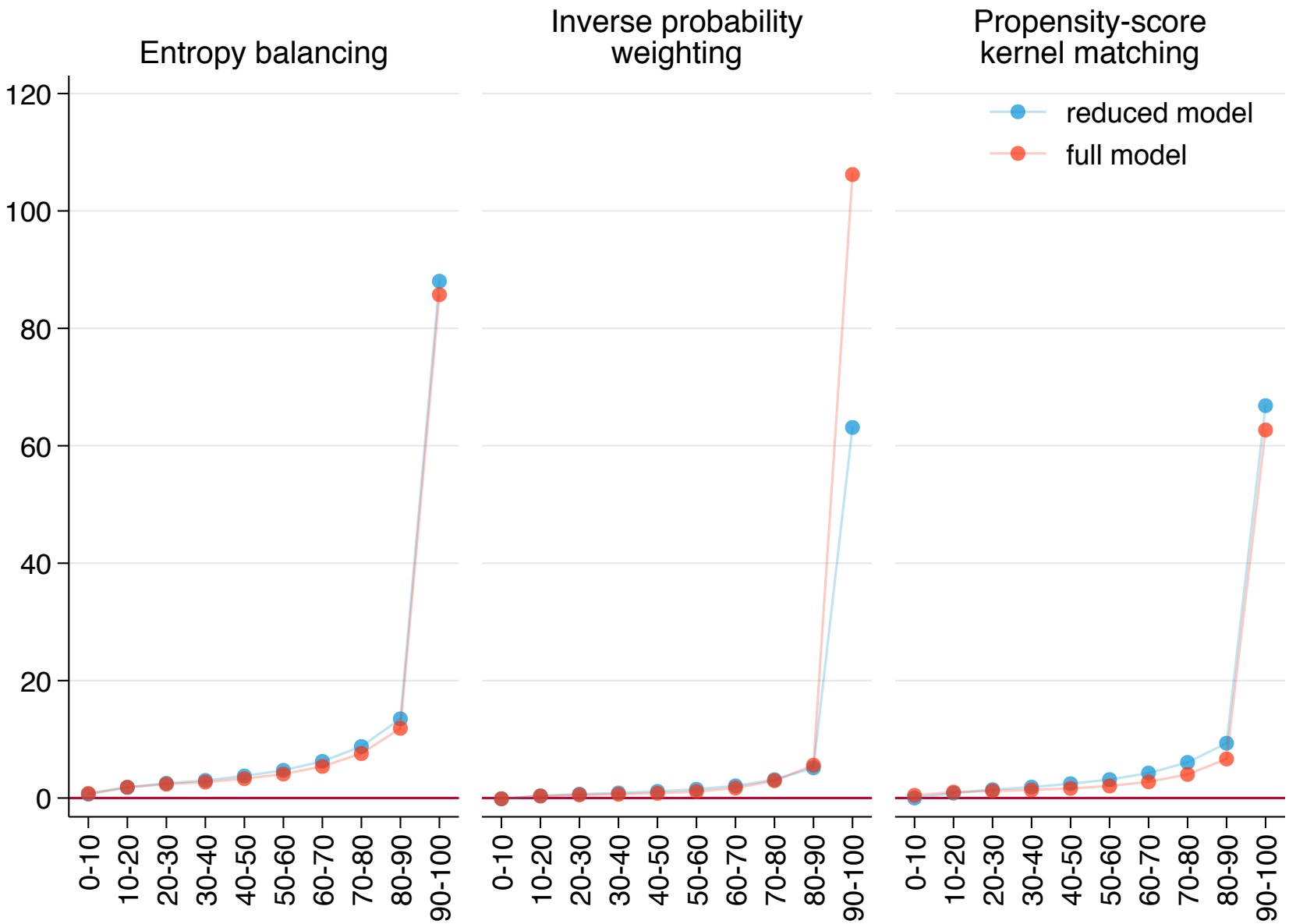
- We use various methods to compute the weights
  - Entropy balancing (Hainmüller 2012)
  - Inverse probability weighting
  - Propensity-score kernel matching (Jann 2017)

- Included variables (at the municipality level)
  - FTA indicators: average (pseudo-equivalized) taxable income, Gini coefficient of (pseudo-equivalized) taxable income
  + population size, distribution of household sizes, age distribution (reduced model)
  + economic structure of work force, proportion of welfare recipients (full model)

Results


Gini of disposable income
Individuals (equivalized)
Households
Unweighted
Entropy balancing
– reduced model
– full model
Inverse probability weighting
– reduced model
– full model
Propensity-score kernel matching
– reduced model
– full model
0 .05 .1 .15 .2 .25 .3 .35 .4 .45
0 .05 .1 .15 .2 .25 .3 .35 .4 .45

Difference in average equivalized disposable income (in 1000) by percentile group between raw and reweighted data:



1 Introduction

2 Data linkage

3 Some preliminary results

4 Reweighting

5 Conclusions

Conclusions

- Our results indicate that Swiss income inequality estimates based on surveys might be substantially biased. This seems mostly due to strong underrepresentation of high income households.
- For example, the Swiss Gini coefficient of equivalized disposable income based on the reweighted cantonal admin data is substantially higher than suggested by comparable survey based analyses (over .40 compared to less than .30).
- Although not shown in this presentation, the admin data, especially if combined with large-scale survey data (which typically lacks information on income, at least in Switzerland), provides excellent opportunities for very fine grained analyses of inequality structures.
- However, there is still a lot to do and many details need to be addressed ...

References

- Bundesamt für Statistik (2013). BFS Aktuell – Einkommen der privaten Haushalte: Einkommensungleichheit wird durch Umverteilung deutlich verringert. Neuchâtel: BFS.
- Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis (Winter 2012) 20 (1), 25-46.
- Jann, B. (2017). KMATCH: Stata module for multivariate-distance and propensity-score matching. Statistical Software Components S458346, Boston College Department of Economics.
- Martinez, I. (2017), Die Topeinkommen in der Schweiz seit 1980: Verteilung und Mobilität. Social Change in Switzerland No. 11. Retrieved from `http://socialchangeswitzerland.ch`
- OECD. (2008). Growing Unequal? Income Distribution and Poverty in OECD Countries. Paris: OECD Publishing.
- OECD (2011). Divided We Stand. Why Inequality Keeps Rising. Paris: OECD Publishing.
- OECD. (2015). In It Together: Why Less Inequality Benefits All. Paris: OECD Publishing.
- Salverda, W., Nolan, B., Checchi, D., Marx, I., McKnight, A. & Toth, I. G. (2014). Changing Inequalities in Rich Countries: Analytical and Comparative Perspectives. Oxford: Oxford University Press.